



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/820,401	03/29/2001	Michael Kahn	MATP-601US	5495
23122	7590	12/11/2003	EXAMINER	
RATNERPRESTIA P O BOX 980 VALLEY FORGE, PA 19482-0980			WOZNIAK, JAMES S	
			ART UNIT	PAPER NUMBER
			2655	

DATE MAILED: 12/11/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

**Office Action Summary**

Application No.

09/820,401

Applicant(s)

KAHN, MICHAEL

Examiner

James S. Wozniak

Art Unit

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 03/29/2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-20 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☒ Claim(s) 7-10 is/are allowed.
- 6) ☒ Claim(s) 1-5, 11-14, 16-20 is/are rejected.
- 7) ☒ Claim(s) 6, 15 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☒ The specification is objected to by the Examiner.
- 10) ☒ The drawing(s) filed on 03/29/2001 is/are: a) ☒ accepted or b) ☐ objected to by the Examiner.
- Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. §§ 119 and 120**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.
- 13) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application) since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.
- a) ☐ The translation of the foreign language provisional application has been received.
- 14) ☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121 since a specific reference was included in the first sentence of the specification or in an Application Data Sheet. 37 CFR 1.78.

**Attachment(s)**

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☒ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449) Paper No(s) 2,3.
- 4) ☐ Interview Summary (PTO-413) Paper No(s). \_\_\_\_\_.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: \_\_\_\_\_.

**Detailed Action**

***Claim Objections***

1. Claim 1 is objected to because of the following informalities:

- “To” in Claim 1, Line 2 is unnecessary and should be deleted

Appropriate correction is required.

***Claim Rejections - 35 USC § 102***

2. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3. **Claims 1, 2, 4, 16, 17, and 19** are rejected under 35 U.S.C. 102(e) as being anticipated by U.S. Patent Publication: 2001/0025241 by Lange et al.

With respect to **Claims 1 and 16**, Lange discloses:

A method and computer readable carrier (carrier defined as a magnetic or optical disk as per paragraph 43 in the application specification) (*conversion of an audio component of an AV signal into captions implemented using a computer readable medium containing a program, Paragraph 16, Lines 6-10*) including computer program instructions, respectively, for displaying text information corresponding to a speech portion of audio signals of a television program to as a closed caption on an video display device (*AV captioning system including a display device,*

*Fig. 1 Element 60, and a speech-to-text processor, Paragraph 16, Lines 3-6, also, the display device can be a television capable of processing closed captioning data, Paragraph 23, Lines 1-5), the method comprising the steps of:*

*Decoding the audio signals of the television program (signal separation processing system, Fig. 1, Element 30, for separating the audio signal from an AV signal, Paragraph 19, Lines 1-7);*

*Filtering the audio signals to extract the speech portion (further processing of an audio signal wherein the speech portion of the audio data is used for conversion to text, which suggests a filtering process for removal of speech from the audio data, Paragraph 27, Lines 1-4);*

*Parsing the speech portion into discrete speech components in accordance with a speech model and grouping the parsed speech components (vocabulary used for recognizing speech as word units, Paragraph 31, Lines 5-7);*

*Identifying words in a database corresponding to the grouped speech components (vocabulary database used for word recognition within a speech component of an audio signal, Paragraph 31, Lines 5-7); and*

*Converting the identified words into text data for display on the display device as the closed caption (text box area for displaying text as it is transcribed, Paragraph 31, Lines 24-25).*

With respect to **Claims 2 and 17**, Lange teaches:

A method and computer readable carrier, according to claims 1 and 16, respectively, wherein the step of filtering the audio signals is performed concurrently with the inherently concurrent step of decoding of later-occurring audio signals of the television program and step of

parsing of earlier occurring speech signals of the television program (*captioning of an AV signal as it is received, Paragraph 27*).

With respect to **Claims 4 and 19**, Lange recites:

A method and computer readable carrier according to claims 1 and 16, respectively, further including the step of formatting the text data into lines of text data for display in a closed caption area of the display device (*suggested option of caption formatting in a two or three-line roll-up format, Paragraph 23, Lines 17-25*).

Thus Lange anticipates the invention as recited in Claims 1, 2, 4, 16, 17, and 19.

### ***Claim Rejections - 35 USC § 103***

4. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

5. **Claims 3 and 18** are rejected under 35 U.S.C. 103(a) as being unpatentable over Lange in view of U.S. Patent: 6,332,122 to Ortega et al.

With respect to **Claims 3 and 18**:

Lange teaches the method and computer program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a

vocabulary database as applied to Claim 1, however Lange does not teach: the use of a speaker independent model in individual word recognition as recited in Claims 3 and 18.

Ortega discloses:

A method and computer readable medium according to claims 1 and 16, respectively, wherein the step of parsing the speech portion into discrete speech components includes the step of employing a speaker independent model to provide individual words as the parsed speech components (*speech frames identified through the use of a speaker independent model, Col. 4, Lines 35-48*).

Lange and Ortega are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a speaker independent model in speech recognition as taught by Ortega with the captioning system and computer program utilizing a vocabulary for word identification as taught by Lange to create a captioning method in which the speaker can be easily identified by a viewer in the case where the system has not been trained for that particular speaker. Therefore, it would have been obvious to combine Ortega with Lange for the benefit of obtaining a captioning method capable of identifying a spoken utterance even if the speech recognition system has not been trained with that speaker's voice, to obtain the invention as specified in Claims 3 and 18.

6. **Claims 5, 11, and 20** are rejected under 35 U.S.C. 103(a) as being unpatentable over Lange in view of U.S. Patent: 6,415,256 to Ditzik.

With respect to **Claims 5 and 20**:

Lange teaches the method and computer program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a vocabulary database as applied to Claim 1, however Lange does not teach: a speaker dependent model for providing individual phonemes extracted from the speech signal as recited in Claim 5.

Ditzik suggests:

A method and computer readable carrier according to claims 1 and 16, respectively, wherein the step of parsing the speech portion into discrete speech components includes the step of employing a speaker dependent model to provide phonemes as the parsed speech components *(speech recognition system designed for speaker dependent or speaker independent operation, Col. 3, Lines 17-19 and phoneme modeling database used for speech recognition that may contain special or generalized vocabularies, Col. 3, Lines 52-59)*.

Lange and Ditzik are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a speaker dependent model in speech recognition to provide phonemes from speech data as taught by Ditzik with the captioning system and computer program utilizing a vocabulary for word identification as taught by Lange to create a captioning method in which the speaker can be easily identified by a viewer in the case where the system has been trained for that particular speaker and in which a best word choice can be made on an individual phoneme basis. Therefore, it would have been obvious to combine Ditzik with Lange for the benefit of obtaining a captioning method capable of identifying a speaker that has been previously trained within a speech recognition system

featuring a more accurate word identification method on a phoneme-by-phoneme basis, to obtain the invention as specified in Claims 5 and 20.

With respect to **Claim 11**:

Lange teaches the method, system, and computer program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a vocabulary database and features decoding, filtering, and processing means in producing a caption as applied to Claim 1, however Lange does not teach: a phoneme identification as recited in Claim 11.

Ditzik discloses:

A phoneme generator, which parses the speech portion into phonemes in accordance with a speech model (*speech processing utilizing a phoneme modeling function, which divides speech into phonemes using phoneme modeling and HMM algorithms, Col. 3, Lines 30-43*);

A database of words, each word being identified as corresponding to a discrete set of phonemes (*phoneme modeling database that contains a grammar rules database for discerning allowable word sequences, word dictionaries, and a phonological rules database for providing data to the phoneme modeling program, Col. 3, Lines 52-67*); and

A word matcher which groups the phonemes provided by the phoneme generator and identifies words in the database corresponding to the grouped phonemes (*word dictionary suggests word recognition from input phonemes, and inherently, phonemes would also be grouped to form a word since words can consists of multiple phonemes. Phoneme groupings in word form are then further grouped in a word sequence using a grammar dictionary, Col. 3, Lines 55-59*).



Lange and Ditzik are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine a means of separating input speech into phonemes for word recognition via a database as taught by Ditzik with the method, system, and computer program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a vocabulary database and features decoding, filtering, and processing means in producing a caption as taught by Lange to create a captioning method in which a best word choice can be made on an individual phoneme basis through phoneme grouping and utilizing a known word database for efficient processing. Therefore, it would have been obvious to combine Ditzik with Lange for the benefit of obtaining a captioning method featuring a more efficient word identification method on a phoneme-by-phoneme basis by examining only the best word choices within a database, to obtain the invention as specified in Claim 11.

With respect to **Claim 12**:

Lange teaches captioning of an AV signal in real-time, which suggests parallel processing as applied to Claim 2, however Lange does not teach a phoneme generator as recited in Claim 12.

Ditzik discloses the phoneme modeling function as applied to Claim 11.

Lange and Ditzik are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine a means of separating input speech into phonemes for word recognition via a database as taught by Ditzik with the method, system, and computer

program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a vocabulary database and features decoding, filtering, and real-time processing means in producing a caption as taught by Lange to create a captioning method in which a best word choice can be made on an individual phoneme basis through phoneme grouping and utilizing a known word database for efficient processing. Therefore, it would have been obvious to combine Ditzik with Lange for the benefit of obtaining a real-time captioning method featuring a more efficient word identification method on a phoneme-by-phoneme basis by examining only the best word choices within a database, to obtain the invention as specified in Claim 12.

With respect to **Claim 14**,

Lange teaches the method, system, and computer program for converting the audio portion of an AV signal into captions that recognizes individual words within speech data by using a vocabulary database and features decoding, filtering, and processing means in producing a caption as applied to Claim 1, however Lange does not teach a speaker dependent speech recognition system containing a phoneme generator as recited in Claim 14.

Ditzik discloses speaker dependent speech recognition as applied to Claim 5 and a phoneme modeling function as applied to Claim 11.

Lange and Ditzik are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a speaker dependent model in speaker dependent speech recognition to provide phonemes from speech data as taught by Ditzik with the captioning system and computer program utilizing a vocabulary for word identification as taught

by Lange to create a captioning method in which the speaker can be easily identified by a viewer in the case where the system has been trained for that particular speaker and in which a best word choice can be made on an individual phoneme basis. Therefore, it would have been obvious to combine Ditzik with Lange for the benefit of obtaining a captioning method capable of identifying a speaker that has been previously trained within a speech recognition system featuring a more accurate word identification method on a phoneme-by-phoneme basis, to obtain the invention as specified in Claim 14.

7. **Claim 13** is rejected under 35 U.S.C. 103(a) as being unpatentable over Lange in view of Ditzik, and in further view of Ortega.

With respect to **Claim 13**:

Lang in view of Ditzik teaches the captioning system and method as applied to Claim 11, but do not teach the use of a speaker independent model as recited in Claim 13.

Ortega recites the use of a speaker independent model in speech frame recognition as applied to Claim 3.

Lange, Ditzik, and Ortega are analogous art because they are from a similar field of endeavor in transcribing text from speech. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to combine the use of a speaker independent model in speech recognition as taught by Ortega with the captioning system and computer program utilizing a vocabulary for word identification from phoneme groupings as taught by Lange in view of Ditzik to create a captioning method in which the speaker can be easily identified by a viewer in the case where the system has not been trained for that particular speaker and in which a best word choice can be made on an individual phoneme basis through

phoneme grouping utilizing a known word database for efficient processing. Therefore, it would have been obvious to combine Ortega with Lange and Ditzik for the benefit of obtaining a more flexible captioning method capable of identifying speech by a speaker even if the speech recognition system has not been trained with that speaker's voice, to obtain the invention as specified in Claim 13.

*Allowable Subject Matter*

8. **Claims 6 and 15** are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

9. **Claims 7-10** are allowed.

10. The following is a statement of reasons for the indication of allowable subject matter:

Prior art teaches the captioning system that converts the audio portion of an AV signal through individual phoneme recognition and word sequencing and features speaker dependent/independent recognition, decoding, filtering, and processing means as applied to Claims 1-5, 11-14, and 16-20. Additionally, "An Automatic Caption-superimposing System with a New Continuous Speech Recognizer" by Imai, Ando, and Miyasaka teaches a captioning system utilizing speech recognition training for error minimization performed before television signal transmission.

Prior art does not teach:

- A transmitted television signal containing training text used for updating a HMM to be used in phoneme identification as recited in **Claims 6 and 15**.

- A transmitted television signal containing training text used for generating a HMM as recited in **Claim 7**.
- **Claims 8, 9, and 10** further limit their parent claims, and thus, are allowable.

### *Conclusion*

11. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

- U.S. Patent: 6,505,153 to Van Thong et al- teaches a method of closed caption production featuring an audio classifier for segmenting speech input, utilizing an HMM for speech recognition.
- U.S. Patent: 6,172,675 to Ahmad et al- teaches a method of audiovisual data manipulation that uses an HMM network containing a phoneme library in the transcription of speech to text.
- "Improving Acoustic Models with Captioned Multimedia Speech" by Jang and Hauptmann- teaches improved acoustic model training as applied to captioning.

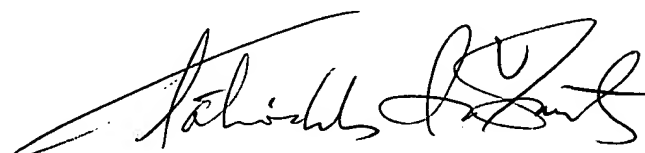
12. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (703) 305-8669 and email is Jwozniak@uspto.gov. The examiner can normally be reached on Mondays-Fridays, 8:30-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Ivars Smits can be reached at (703) 306-3011. The fax/phone number for the Technology Center 2600 where this application is assigned is (703) 872-9306.

Art Unit: 2655

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the technology center receptionist whose telephone number is (703) 306-0377.

James S. Wozniak  
12/4/2003

A handwritten signature in black ink, appearing to read 'Tālivaldis Ivars Šmits', written in a cursive style.

TĀLIVALDIS IVARS ŠMITS  
PRIMARY EXAMINER